

Hutch Integrated Data Repository and Archive (HIDRA)

Paul A. Fearn, MBA
Director, Biomedical Informatics
Fred Hutch Cancer Research Center

October 23, 2014













HIDRA

- Vision and Strategy
- Systems & Requirements
- Program Overview
- Argos user interface
- Next steps

Hutch Integrated Data Repository & Archive (HIDRA) Strategic Drivers



Catalyzing our efforts to build strength in clinical molecular diagnostics and precision oncology

- FHCRC "Center News", 4/1/13



Contribute to building a strong Fred Hutchinson /
University of Washington
Cancer Consortium [translational research] program
- Center Strategic Plan 2010-2015



Strengthening the Consortium's clinical research programs and infrastructure to permit more rapid development of diagnostics and therapeutics

- Senior Leadership, Cancer Consortium, RE: areas that require continuing resource commitment

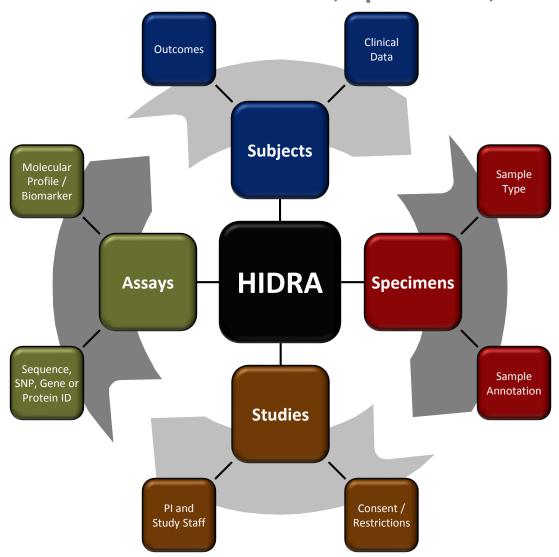


Need an integrated database approach and a Consortium-wide informatics platform strategy

- CCSG Reviewers, 2008

Vision for HIDRA Data Integration Scope & Axes

Master Indexes of Consortium Patients, Specimens, Studies and Assays



HIDRA Goals

Rapid-Learning Informatics Platform for Competitive NCI CCC

- Enable us to learn from every patient who comes through the door, and integrate that knowledge back into the clinical care
 - Use of clinical data for research, activities preparatory to research, healthcare operations, QI/QA, and public health reporting purposes

Integrate

- Integrate data and systems across all disease groups
- Link specimen, genomic and other assay data with clinical data
- Integrate security/permissions with Consortium CTMS

Automate or facilitate manual, repetitive work

- Manual data abstraction, feeds and NLP from medical records
- Outcomes data (e.g. CSS, long-term follow-up and patient reported)

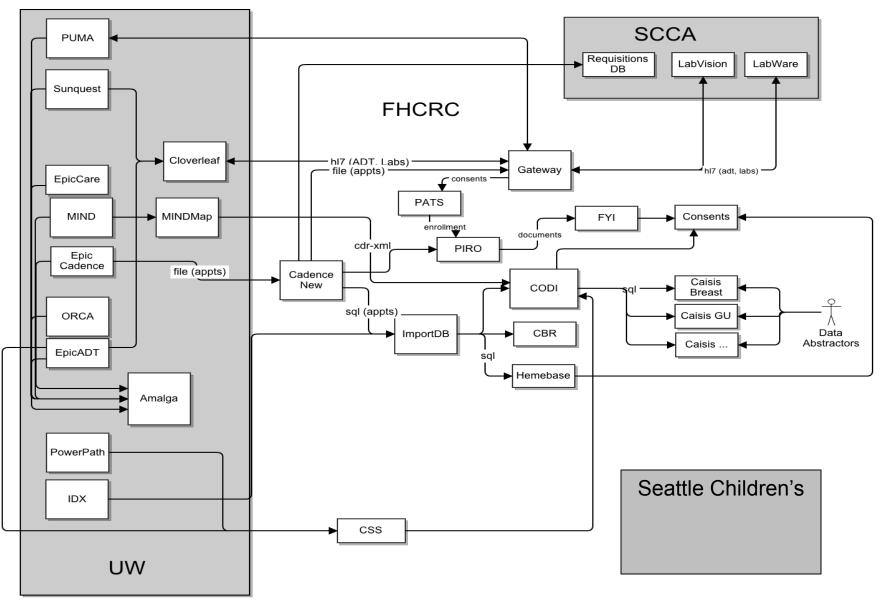
Strong competitive platform

Be ready for FISMA security or FDA regulatory reviews or audits

HIDRA

- Vision and Strategy
- Systems & Requirements
- Program Overview
- Argos user interface
- Next steps

Pre-HIDRA Informatics Ecosystem at Fred Hutch Consortium Data Flows before HIDRA Core



NLP Requirements Analysis

We condensed over 14,000 existing and desired fields from 13 different disease groups into just under 4,000 individual elements

65% of data elements come from unstructured sources

(this estimate was made assuming that patients are diagnosed and receive all of their treatment within the consortium)

15% of data elements could be patient reported

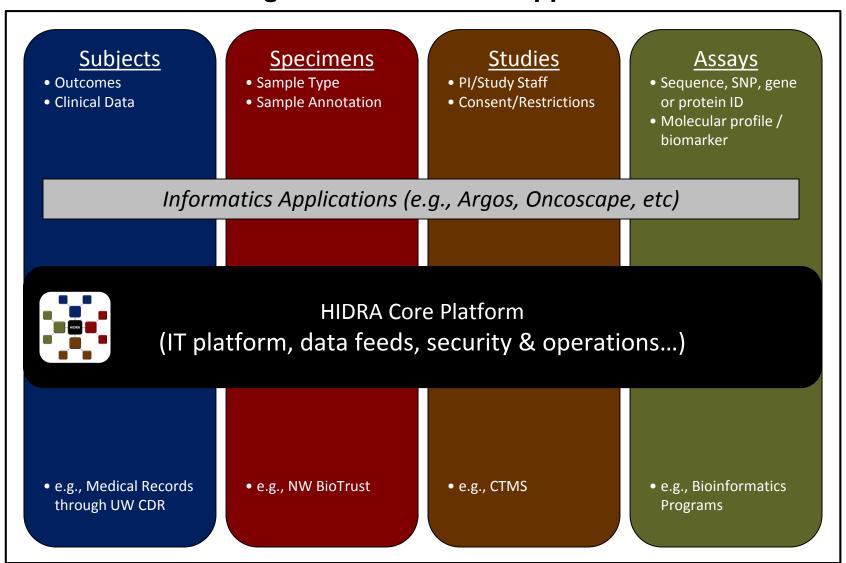
(about 75% of these elements are currently coming from unstructured sources)

15% of data elements are computed from other elements

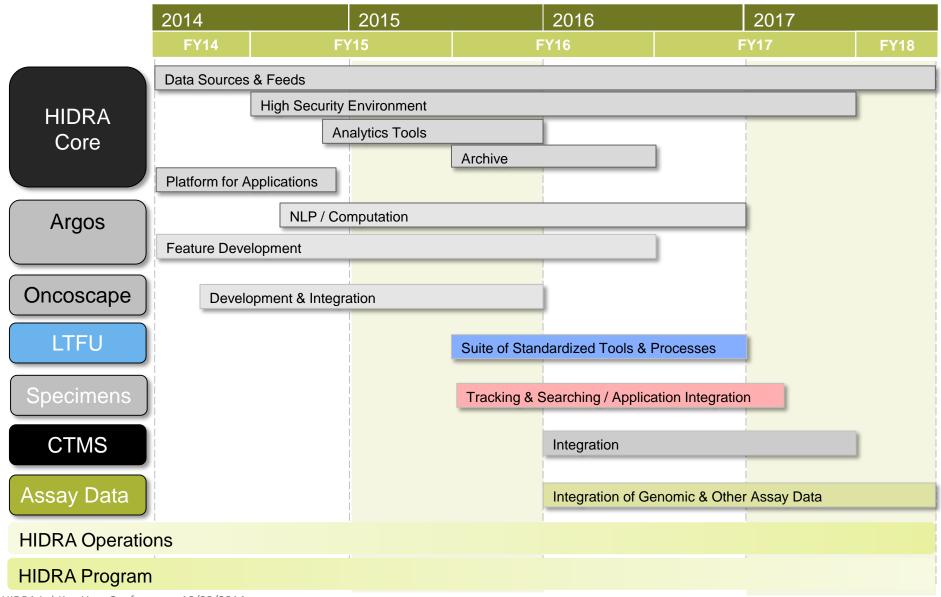
HIDRA

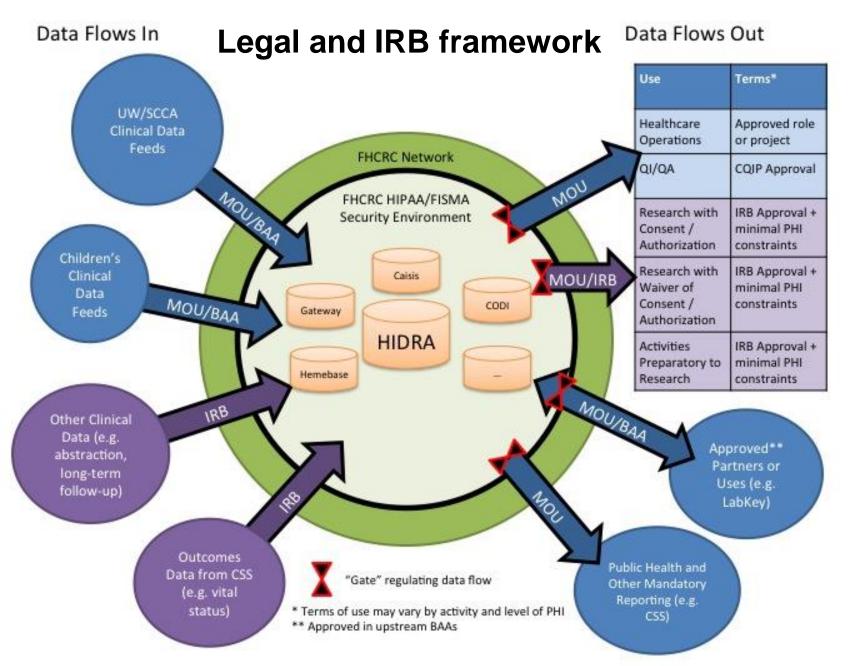
- Vision and Strategy
- Systems & Requirements
- Program Overview
- Argos user interface
- Next steps

HIDRA Program: Platform and Applications Framework

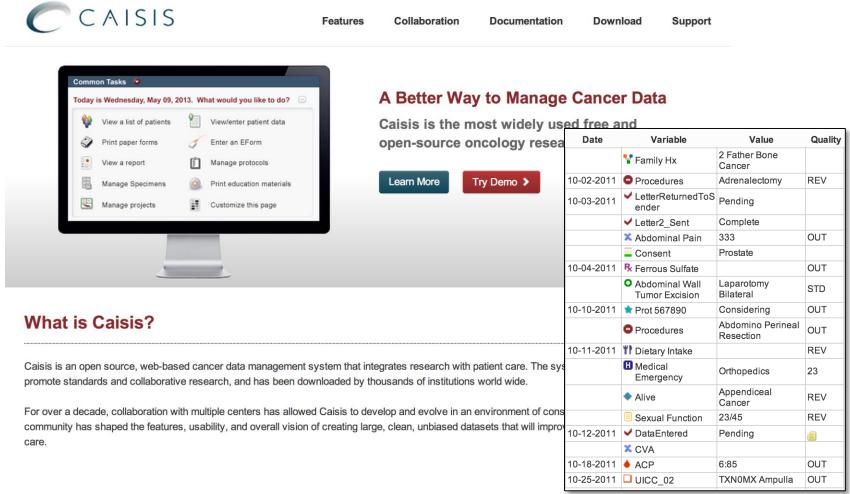


HIDRA Program Road Map

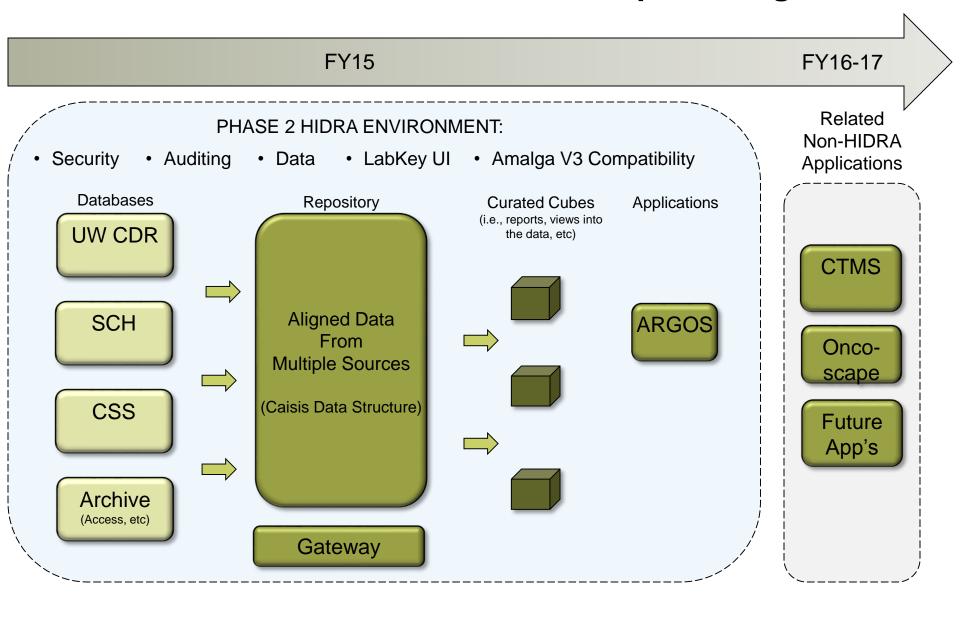




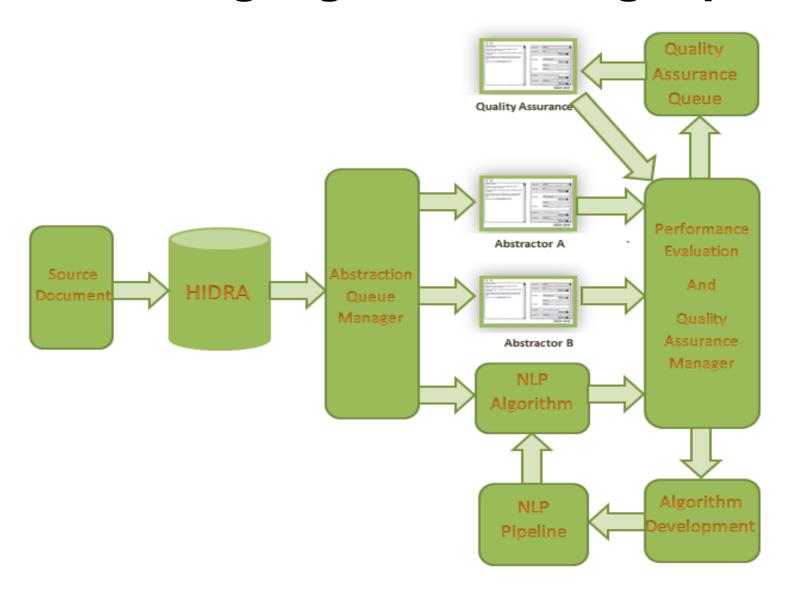
Past Work – Caisis @ MSKCC 1998 - present



HIDRA Environment: Conceptual Diagram



Natural Language Processing Pipeline







mammogram was obtained dated 01/28/12, which showed a mass in the right breast. On 02/10/12, she underwent an ultrasound-guided biopsy. The pathology showed an infiltrating ductal carcinoma Nottingham grade II. The tumor was ER positive, PR positive and HER-2/neu negative. On 02/22/12, she underwent a lumpectomy and sentinel lymph node biopsy. The pathology showed a 3.3 cm infiltrating ductal carcinoma grade I, one sentinel lymph node was negative. Therefore it was a T2, N0, M0 stage IIA breast cancer. Of note, at that time she was taking hormone replacement therapy and that was stopped. She underwent radiation treatment ending in May 2008. She then started on Arimidex, but unfortunately she did not tolerate the Arimidex and I changed her to Femara. She also did not tolerate the Femara and I changed it to tamoxifen. She did not tolerate the tamoxifen and therefore when I saw her on 11/23/12, she decided that she would take no further antiestrogen therapy. She met with me again on 02/22/13, and decided she wants to rechallenge herself with tamoxifen. When I saw her on 04/28/13, she was really doing guite well with tamoxifen. She tells me 2 weeks after that visit, she developed toxicity from the tamoxifen and therefore stopped it herself. She is not going take to any further tamoxifen.

Overall, she is feeling well. She has a good energy level and her ECOG performance status is 0. She denies any fevers, chills, or night sweats. No lymphadenopathy. No nausea or vomiting. No change in bowel or bladder habits.

CURRENT MEDICATIONS: Avapro 300 mg q.d., Pepcid q.d.,

Zyrtec p.r.n., and calcium q.d.

ALLERGIES: Sulfa, Betadine, and IV contrast.

PROCEDURE:	Mammogram
MAMMOGRAM:	01/28/12 Mass on right side
PROCEDURE:	Ultrasound-guided biopsy 02/10/12
TUMOR:	Invasive ductal carcinoma
Endocrine:	ER positive, PR negative
HER2:	HER2/neu negative
Nottingham:	Grade 2
Procedure:	Lumpectomy
	02/22/12
Staging:	T2 Tumor size: 3.3cm
Lymph:	N0

CANCEL

VERIFY

HIDRA

- Vision and Strategy
- Systems & Requirements
- Program Overview
- Argos user interface
- Next steps



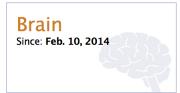
Select your Argos portal.

This sets your filter categories and more. You can change this any time by clicking the Argos logo.

All



By Disease Group

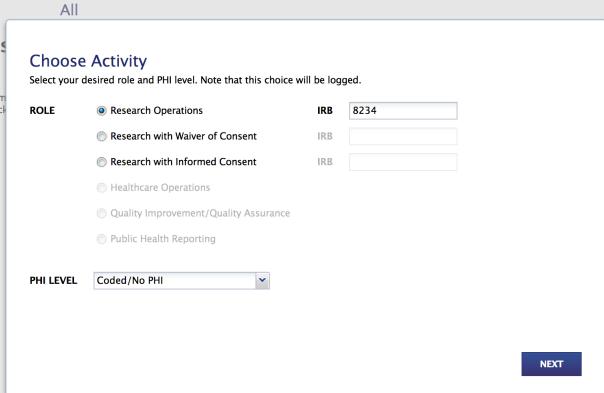


By Study



Select your Argos portal.

This sets your filter categories and m You can change this any time by click Argos logo.





Select your Argos portal.

This sets your filter categories and m You can change this any time by click Argos logo. All

Terms of Use

▼ I AGREE to the terms of use for **Coded/No PHI** in **Research Operations** in the **Brain** group.

- "Confidential Information" means Any data in HIDRA that is linked to or could be used to identify a patient or subject, including 1) electronic data feeds from UW Medicine, SCCA and Children's medical records (Exhibits B.1 and B.2 in HIDRA MOU); 2) other clinical data manually abstracted data from UW Medicine, SCCA and Children's medical records; 3) other clinical data from long-term follow-up that has been manually abstracted from medical records; 4) outcomes data from CSS; 5) data about associated specimens (e.g. from NW BioTrust); 6) data about associated studies (e.g. from CTMS); and 7) data about associated assays (e.g. molecular data associated with the Consortium patients and subjects in HIDRA).
- 2. I agree not to make use of, disseminate, disclose or in any way circulate any Confidential Information except as expressly permitted by this Confidentiality Pledge. Confidential Information may be published or otherwise disclosed in connection with the study entitled "Enrichment, Linkage and Secondary Use of Clinical, Biospecimen and Study Data from Hutch Integrated Data Repository and Archive (HIDRA)." (Institutional Review File #8234) provided, however, that no disclosure may be made which permits identification of any individual patient or the patient's physician unless permitted by applicable law and approved by an Institutional Review Board of FHCRC. Confidential Information may also be disclosed to other persons working on the Study who have signed a Confidentiality Pledge.

CANCEL

OK

Brain Group

Research Operations, IRB 8234 | Coded/No PHI

Overall Patient Statistics

	This Month	This Year	Total
Embryonal Tumors	0	0	9
Ependymal Tumors	0	0	16
Gliomas	0	22	487
Medical Therapy	0	16	314
Procedures	0	8	487
Radiation Therapy	0	4	367
Meningothelial Tumors	0	4	149
Metastatic Tumors	0	0	1
Nerve Sheath Tumors	0	1	148
Other	0	0	40
Other Brain Tumors	0	0	19
Other Neuroepithelial Tumors	0	0	15
Sellar Tumors	0	0	6

DASHBOARD

PATIENTS

SPECIMENS

STUDIES

ASSAYS

REPORTS

1,735

Patients

2,788

Specimens

26 Studies

ACTIVE FILTERS

In Saved Group

(none)

New Filters

(none)

My Links

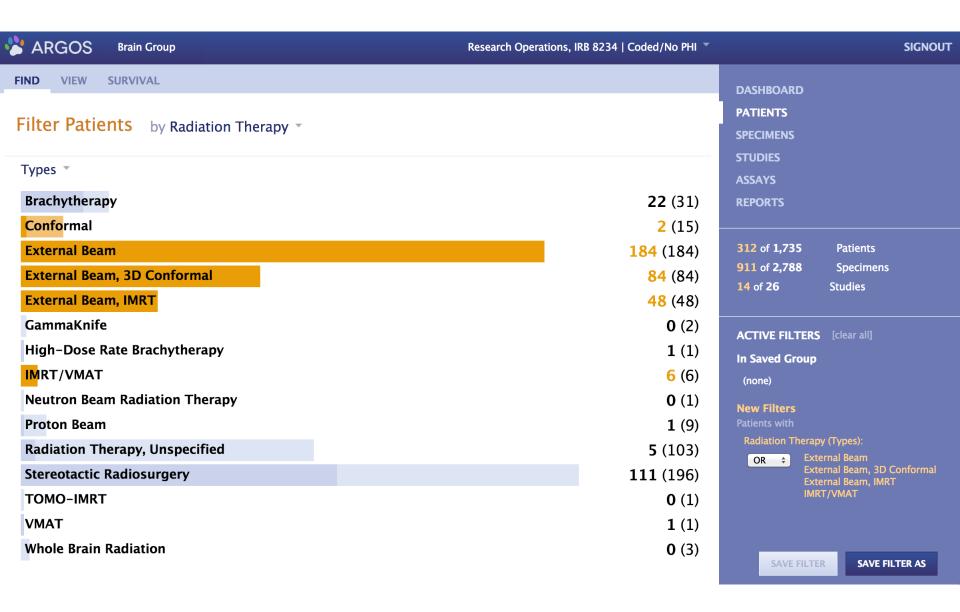
Saved Filter: MyFilter
Saved Filter: study
Saved Filter: new filter!
Saved Filter: DemoMonday
Saved Filter: Alex spence filter

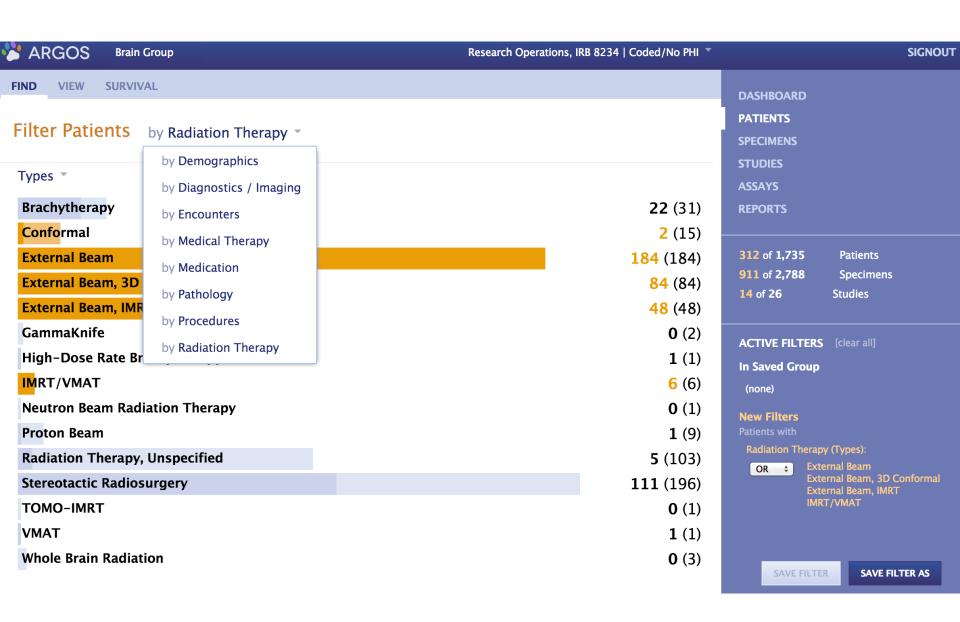
Saved Filter: Donald Born Pathology records

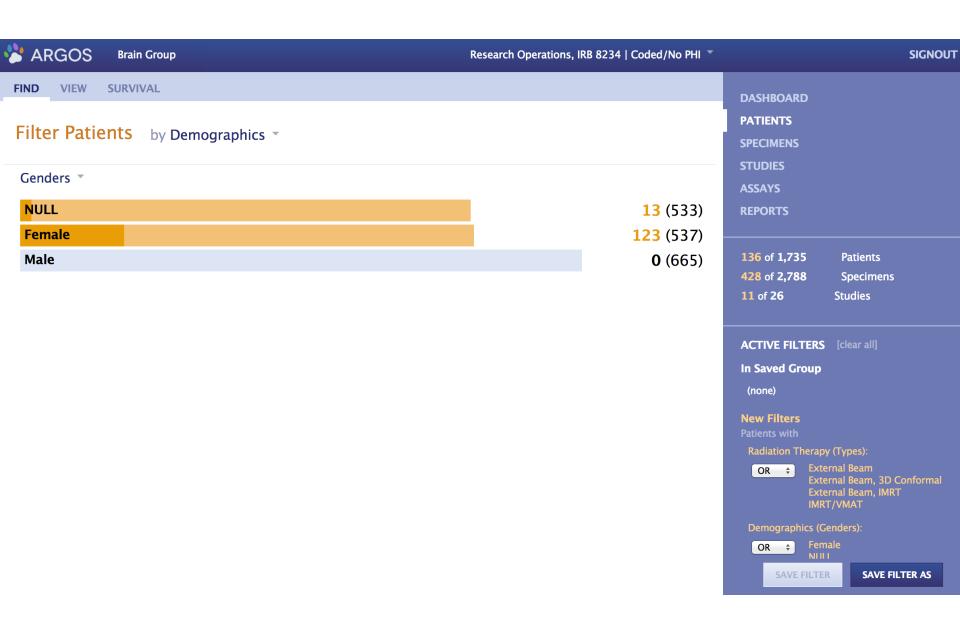
SAVE FILTER

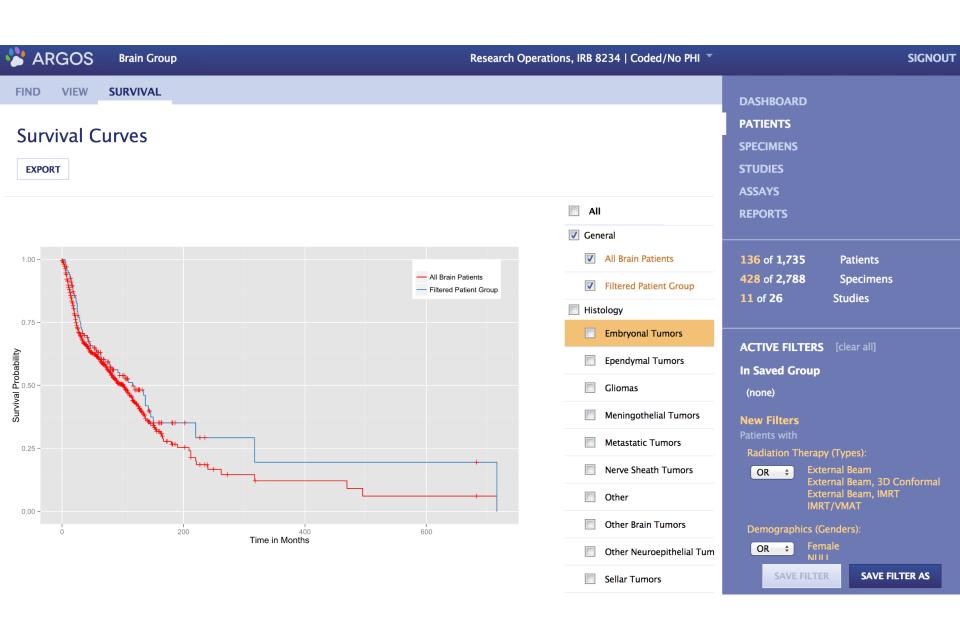
SAVE FILTER AS

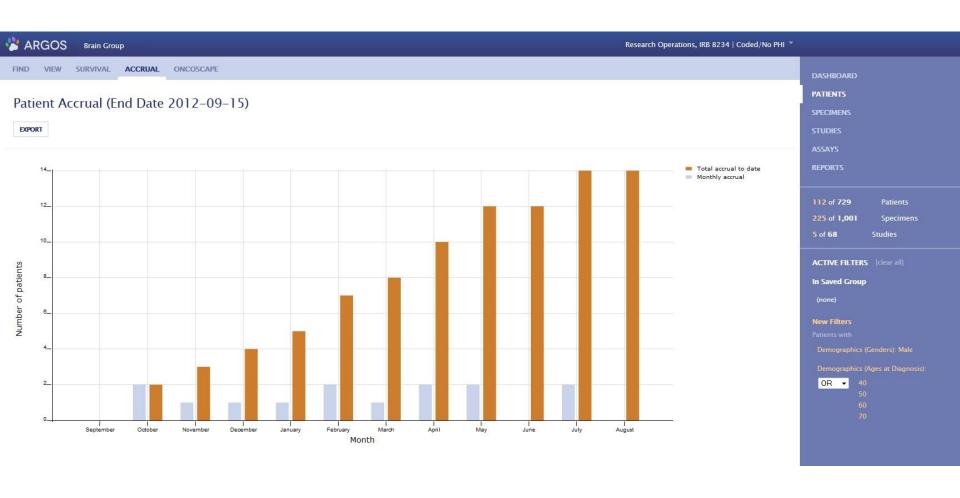
ARGOS Brain Group	Research Operations, IRB 8234 Coded/No PHI	SIGNOUT
Filter Patients		DASHBOARD PATIENTS SPECIMENS STUDIES ASSAYS REPORTS
by Demographics by Diagnostics / Imaging	3 genders, 11 races, 1 ethnicity, 10 ages, 10 ages at diagnosis, 7 years of survival, 10 ages at first surgery 4 types, 4 diseases, 21 results	
by Encounters by Medical Therapy	29 kps, 116 physicians, 10 heights, 32 weights, 12 bsas, 44 bmis 80 agents, 23 years, 13 routes, 45 cycles	1,735 Patients 2,788 Specimens 26 Studies
by Medication by Pathology	0 types, 0 medications 75 histologies, 29 secondary histologies, 744 specimen types, 70 sites, 9 sides, 12 institutions, 39 pathologists, 5 diseases, 14 grades, 35 test results 67 procedures, 51 operating room details/institutions, 128 case surgeons, 31 years, 80 sites, 5	ACTIVE FILTERS In Saved Group (none) New Filters (none)
by Procedures by Radiation Therapy	institutions, 7 services 15 types, 6 diseases, 28 years, 163 sites, 4 isotopes, 163 targets, 80 physicians, 47 institutions	
		SAVE FILTER SAVE FILTER AS

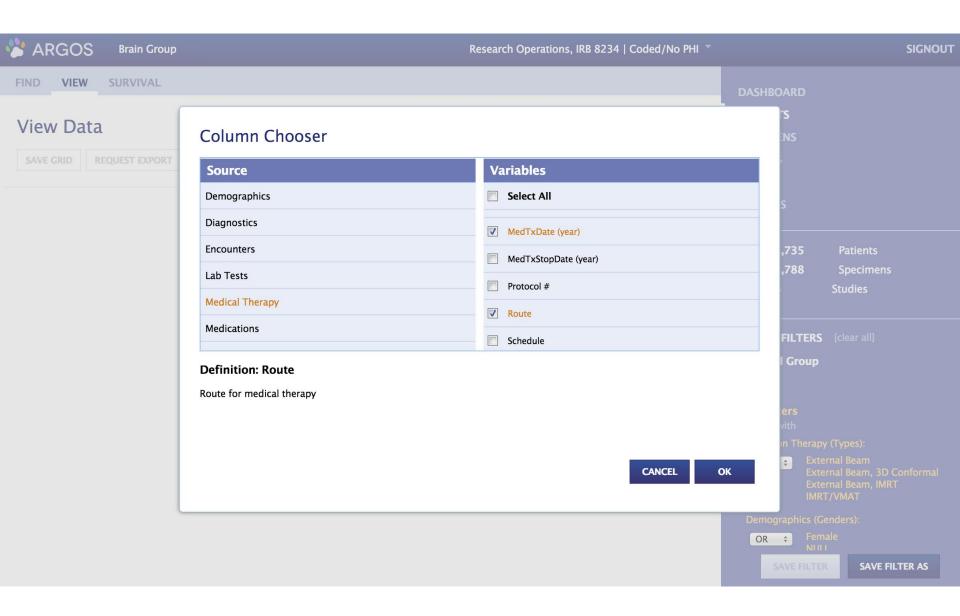


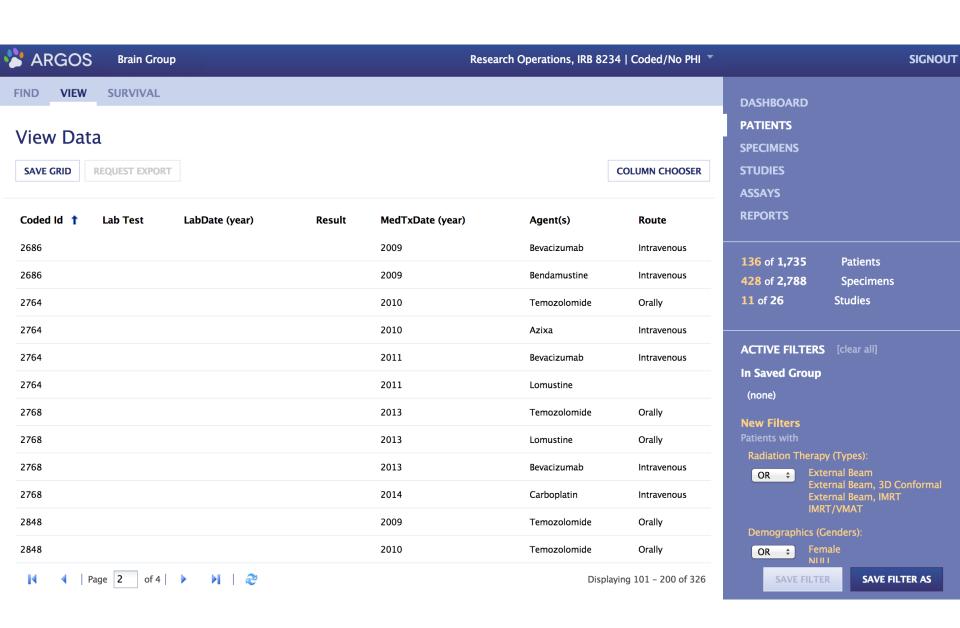


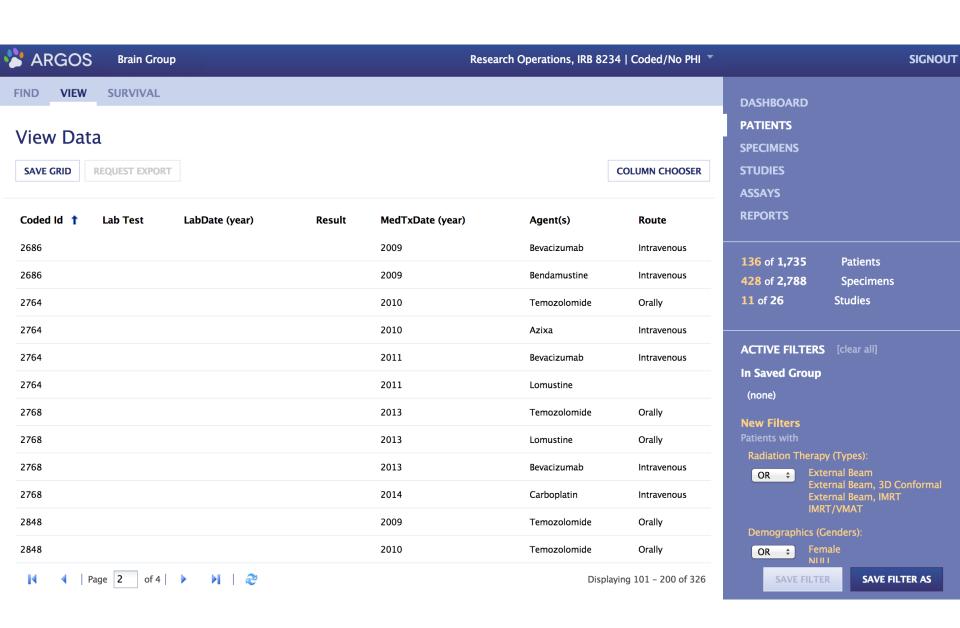


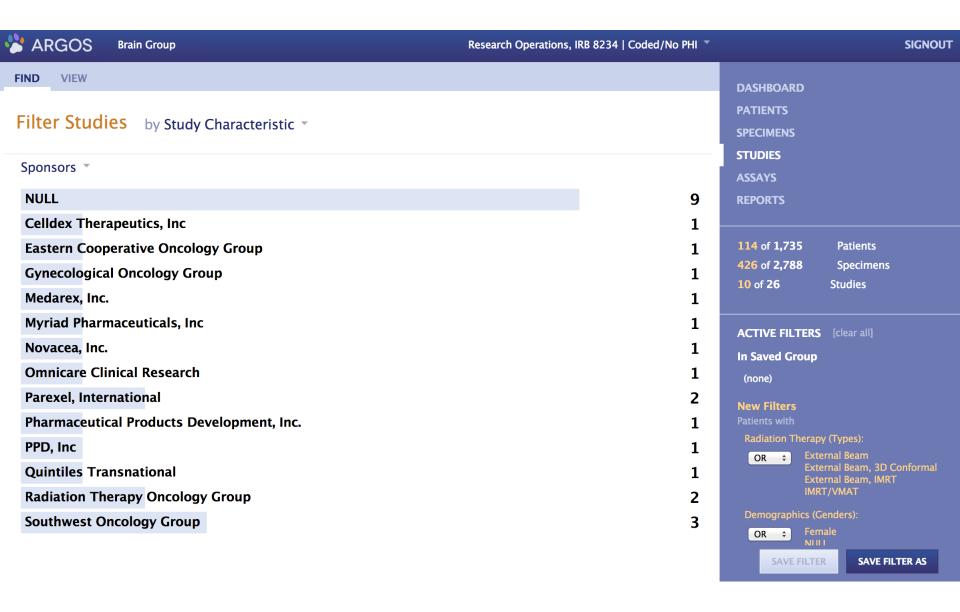








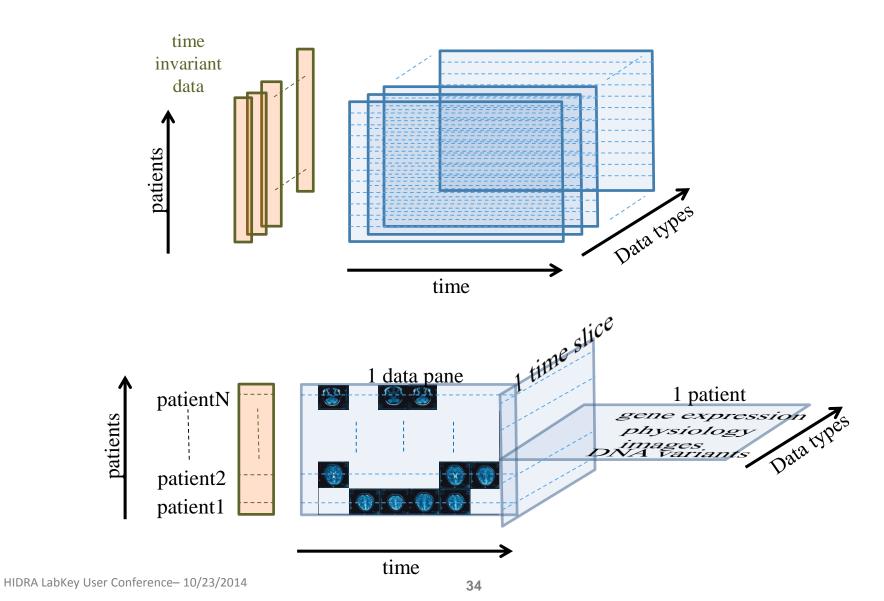




Oncoscape

- Research and Innovation framework for rapid prototyping of features
- Candidate features for migration to Argos on LabKey Server (e.g. patient timelines graph)

a menu-driven, web-based, platform for exploration & analysis of multidimensional clinical data



http://oncoscape.fhcrc.org/

User: Guest

Oncoscape

Markers & Patients

(version 1.1.30, 16 sep 2014)

This website is undergoing frequent modifications.

Use at your own risk.

Note: Oncoscape works best with Chrome Version 37 or higher and a high resolution screen.

PLSR

GBM Pathways

Angiogenesis

Available Data

Oncoscape

← → C

Dashboard

- Table of Contents
- ▼ Features To Come...

ncoscape.fhcrc.org

Clinical Data

Timelines

- Saved Selections
- Interactive Kaplan Meier Plots
- · Expression subtyping tool
- Expression correlation to TCGA samples using MDS
- Tool to find TCGA samples with similar mutational profiles
- Expression clustering and heatmaps
- Gene set enrichment analysis for user-selected groups
- Differential expression analysis for user-selected groups
- · Hallmarks of Cancer

Edit The To-Do Table

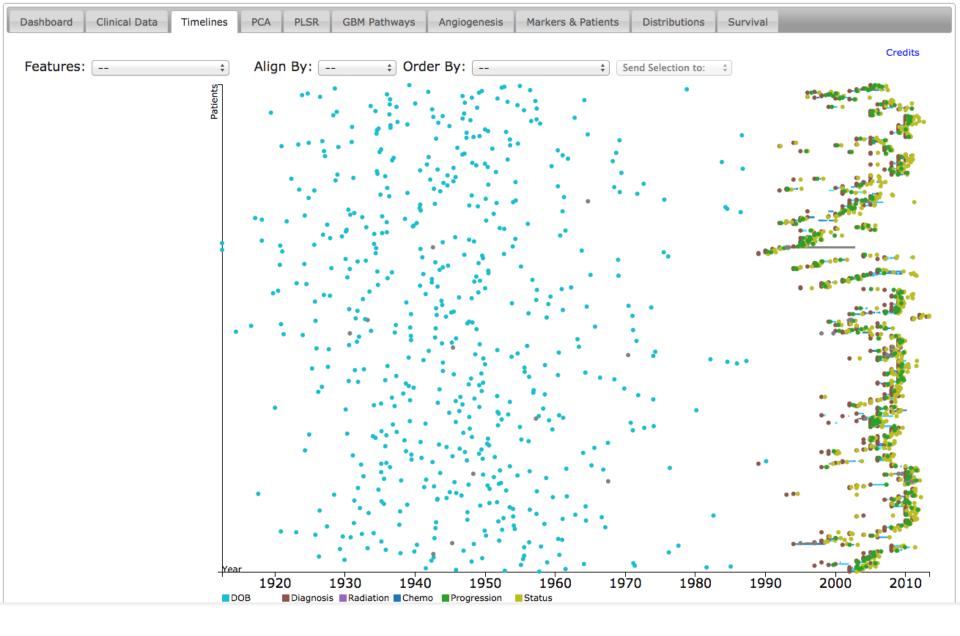
About Oncoscape

Oncoscape is developed at the Fred Hutchinson Cancer Research Center under the auspices of the Solid Tumor Translational Research initiative.

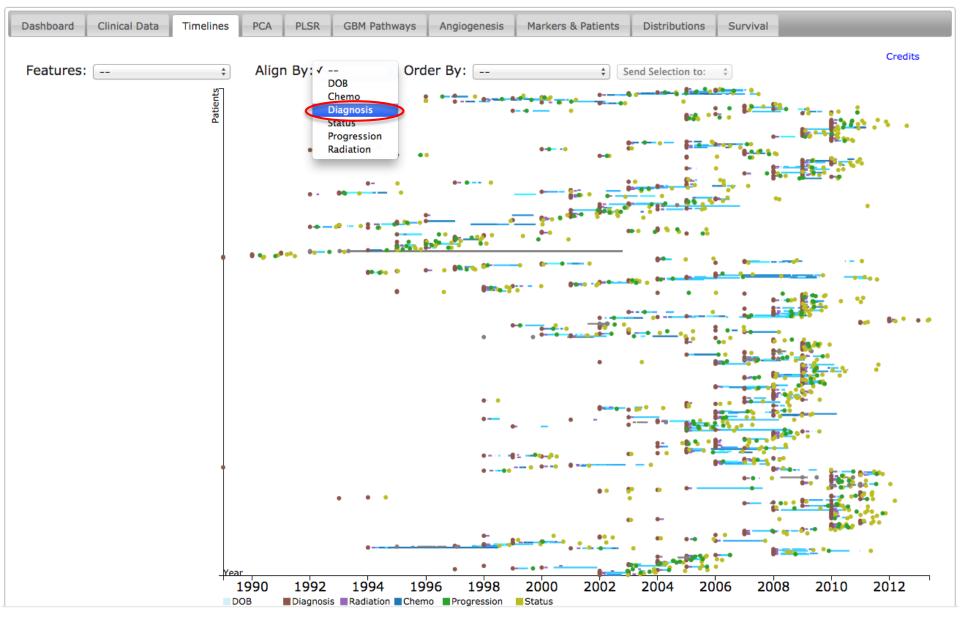
Oncoscape is a web-based, menu-driven analysis and visualization platform for large-scale, heterogeneous clinical and molecular patient timeline data as exemplified by the Fred Hutch HIDRA database.

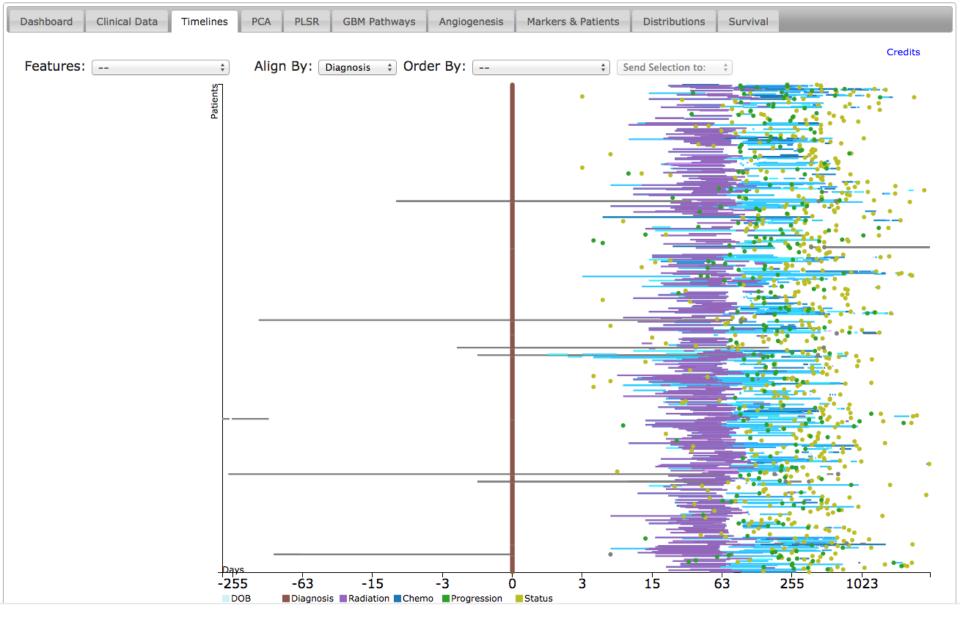
Oncoscape was conceived, and is managed, by a Steering Committee comprising: <u>Eric Holland</u>, <u>Desert Horse-Grant</u>, <u>Paul Fearn</u>, <u>Paul Shannon</u>, <u>Lisa</u> McFerrin, and Hamid Bolouri.

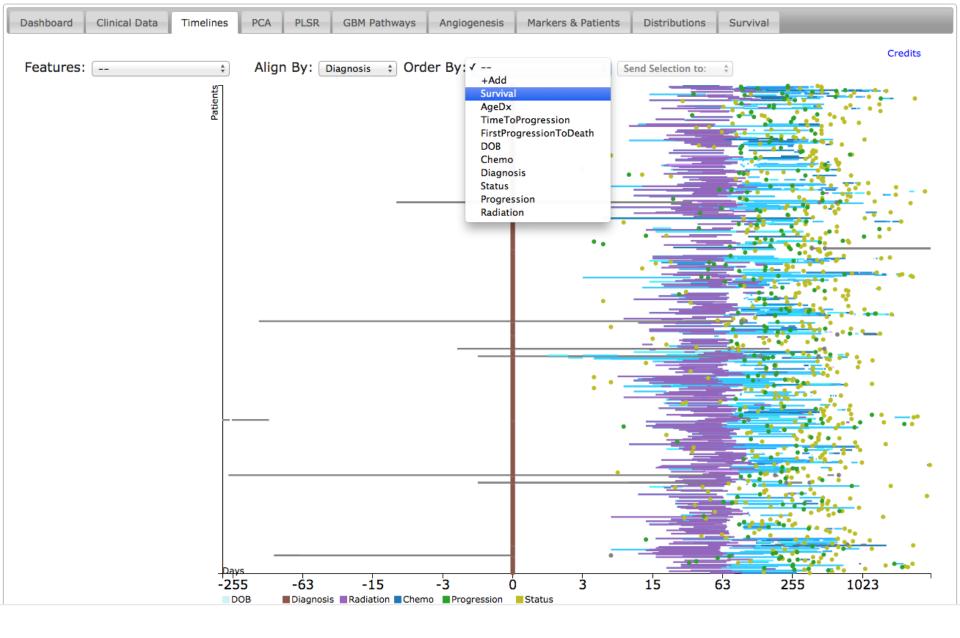
Paul Shannon (lead) and Lisa McFerrin are the primary developers of Oncoscape, with additional code contributions by Cliff Rostomily and Hamid Bolouri.

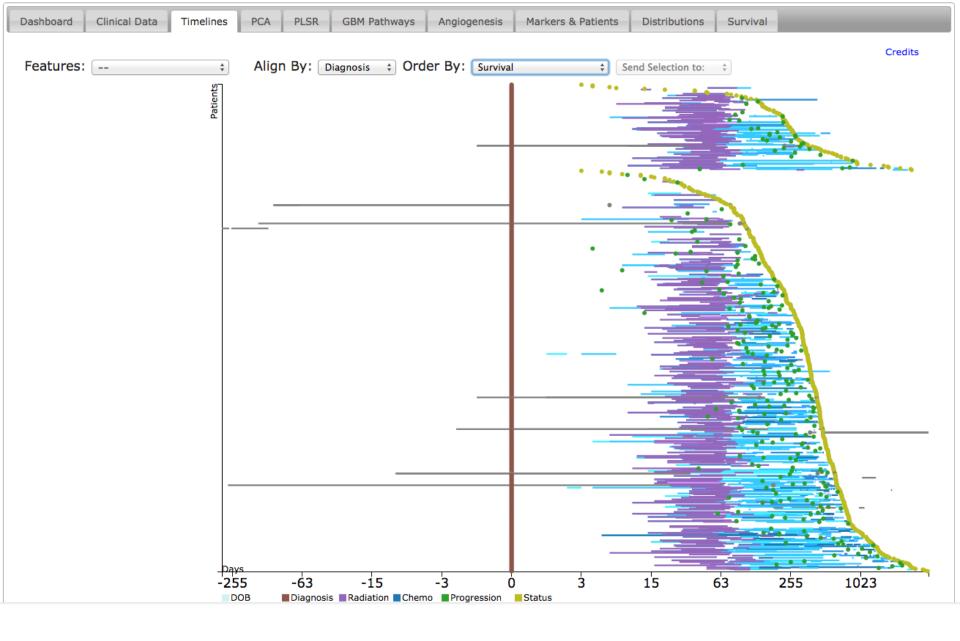


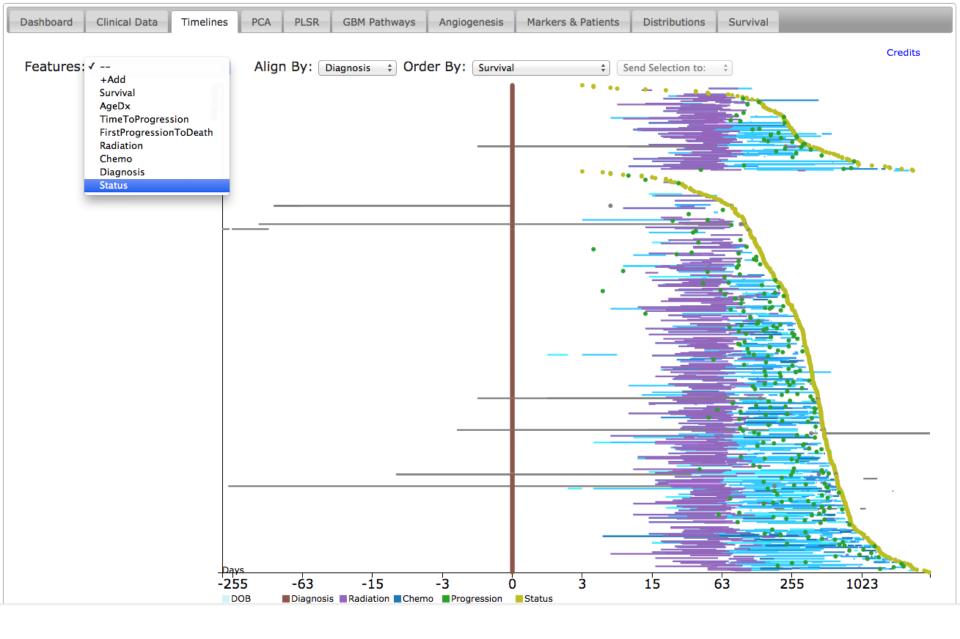


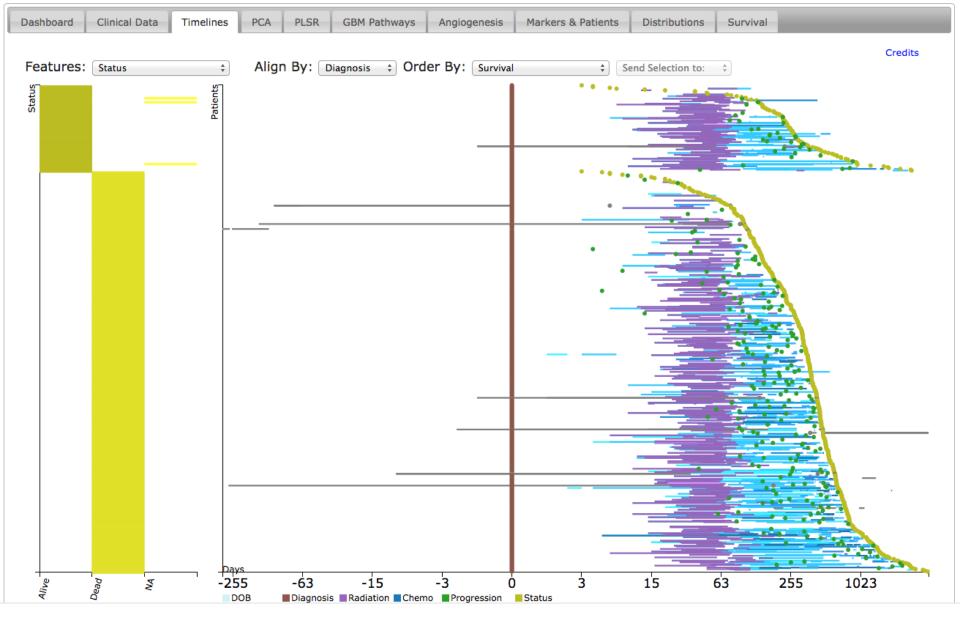


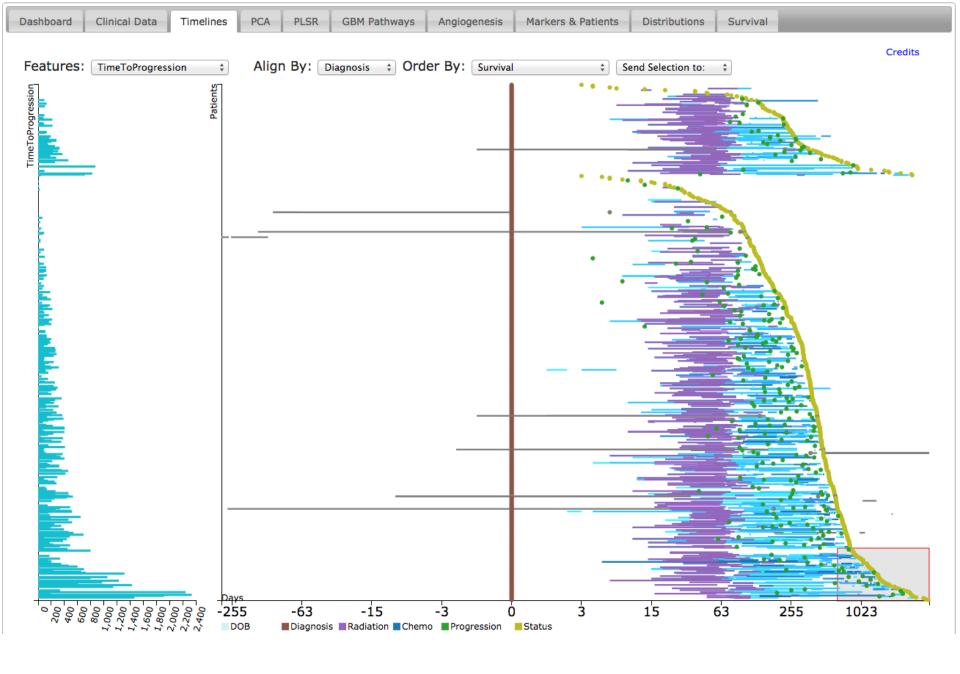


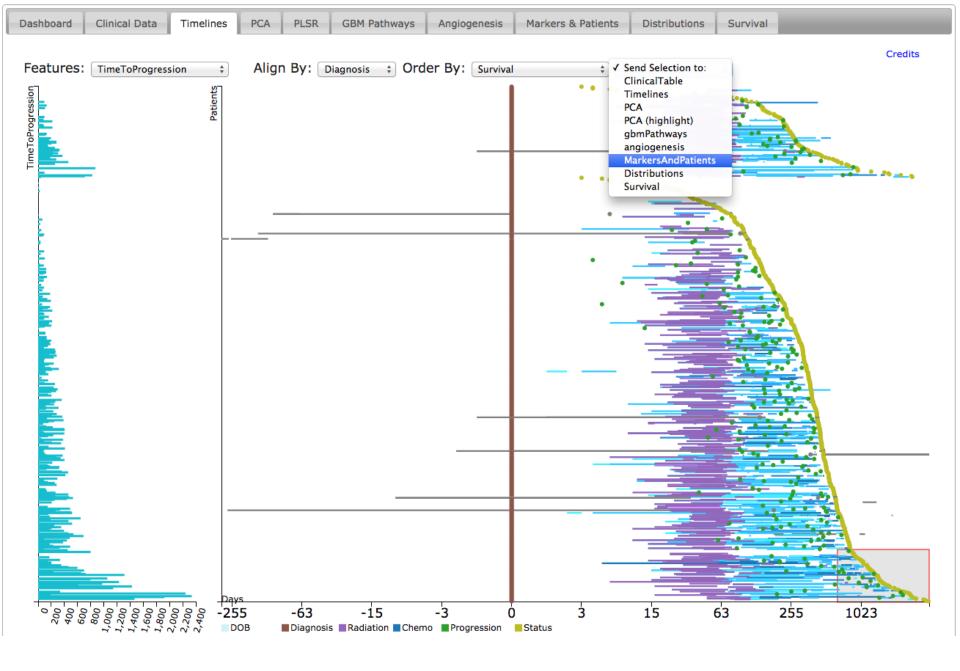


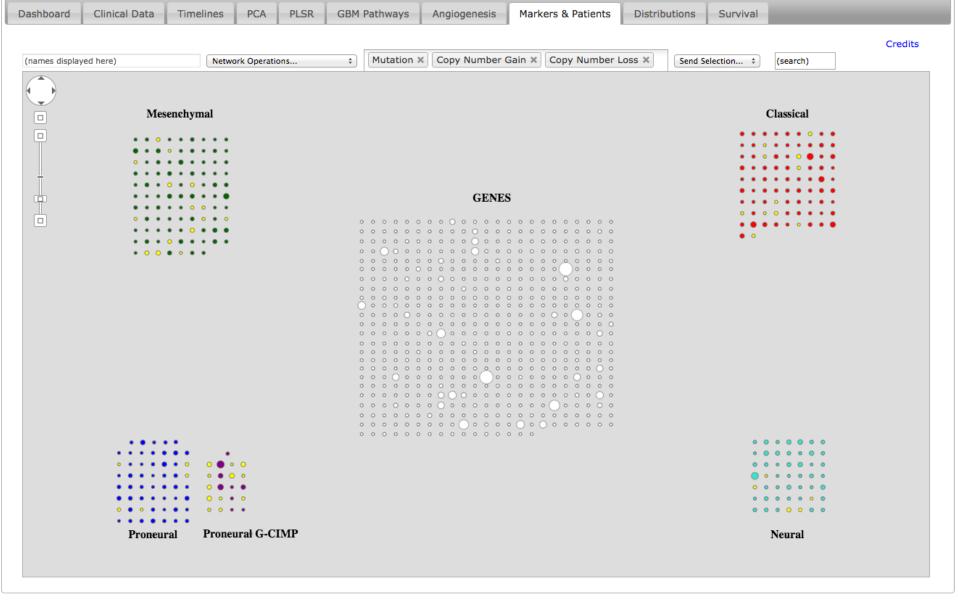












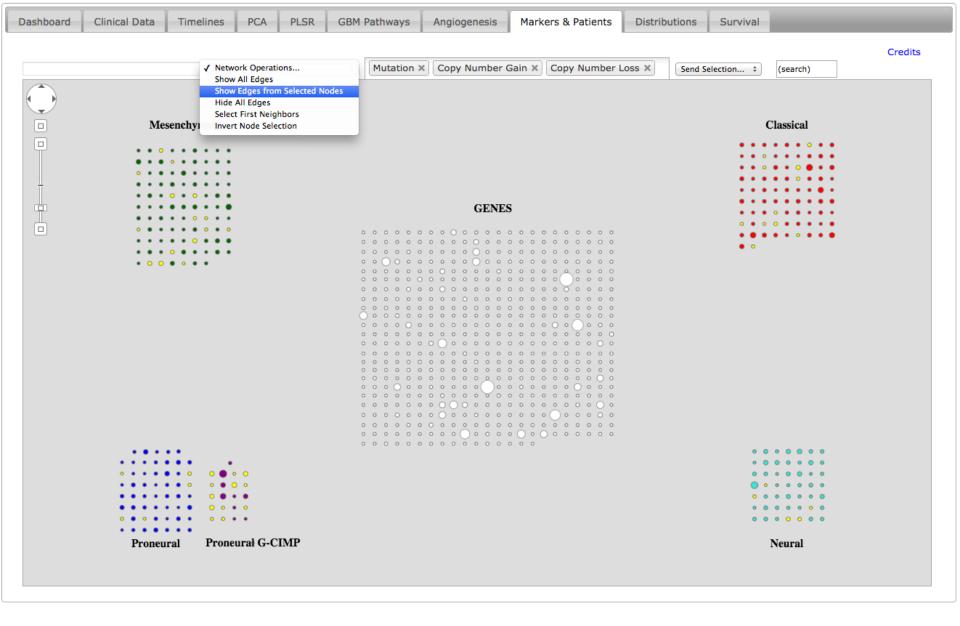
Mesenchymal: 15/97 (15%) Classical: 11/83 (13%)

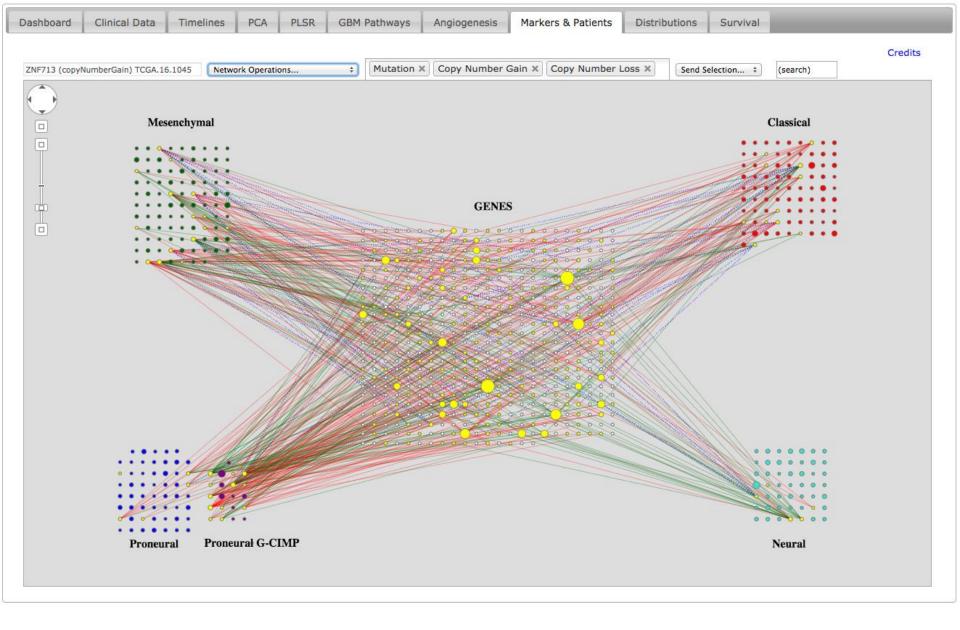
Proneural:

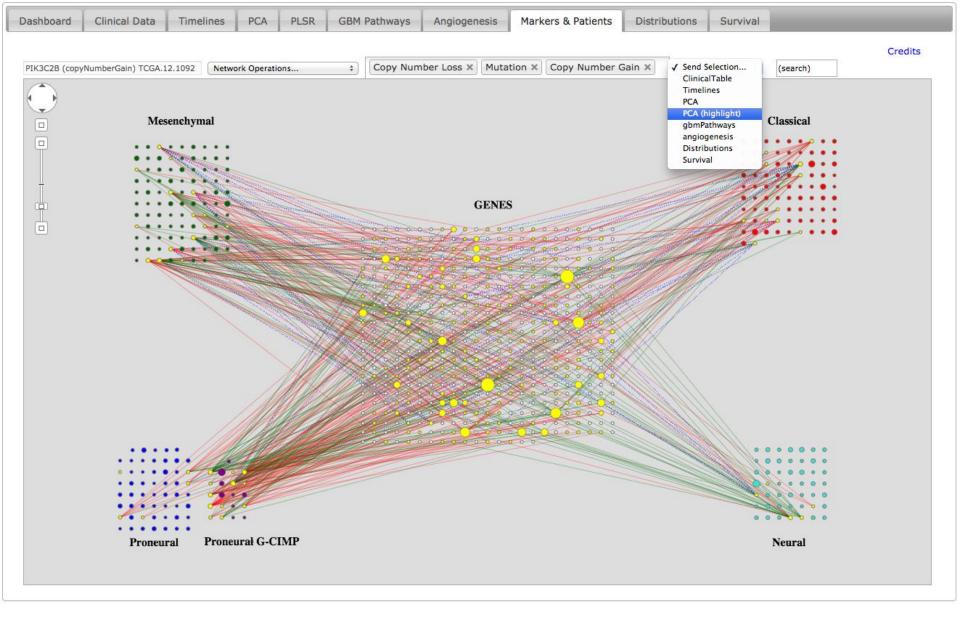
5/54 (9%)

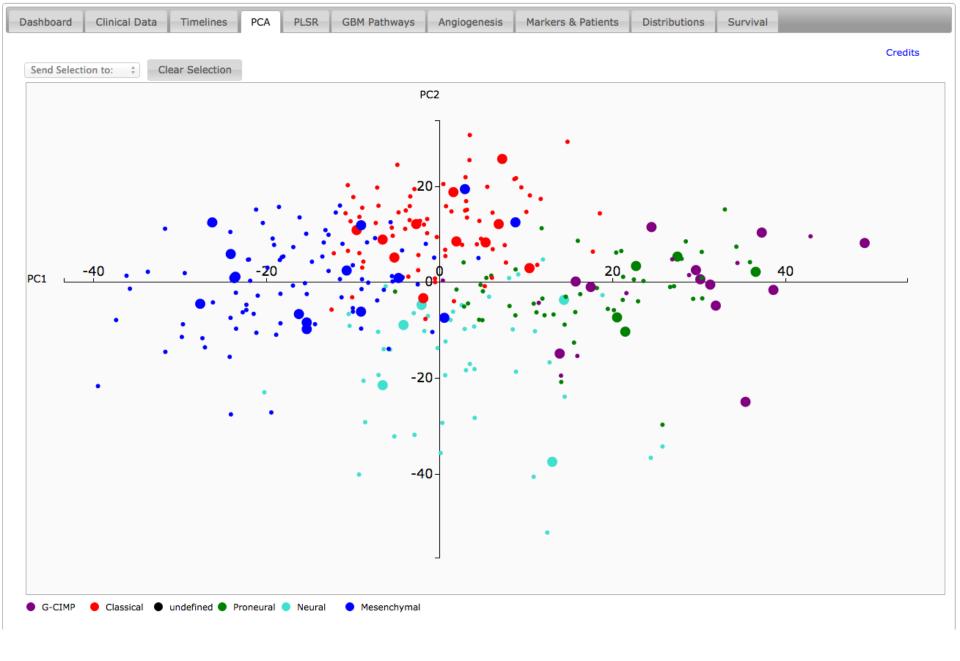
Neural: 5/49 (10%) 12/21 (57%)

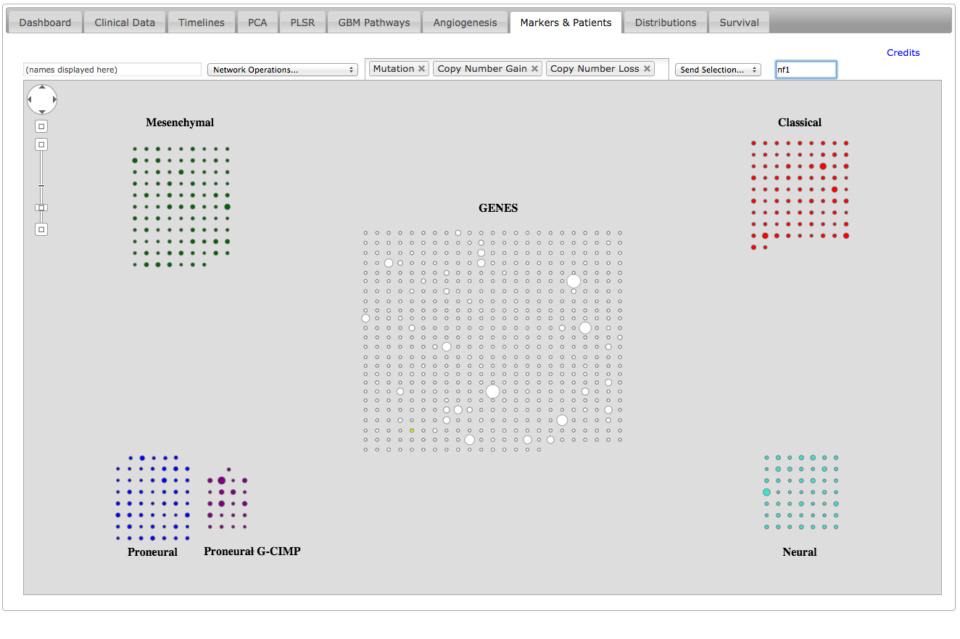
HIDRA LabKey User Conference— 10/23/2014 G-CIMP: 46 12/21 (57%

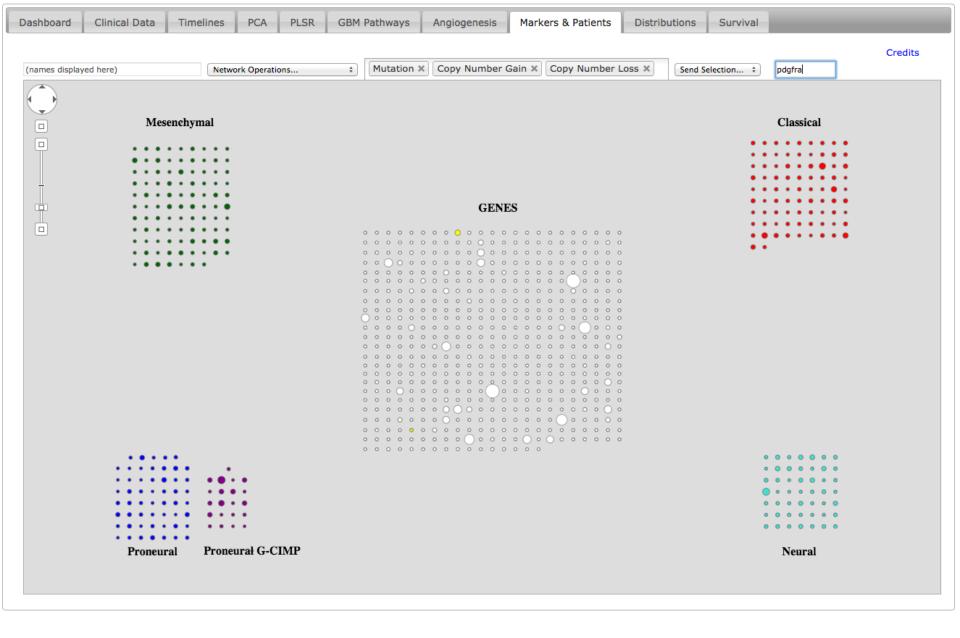


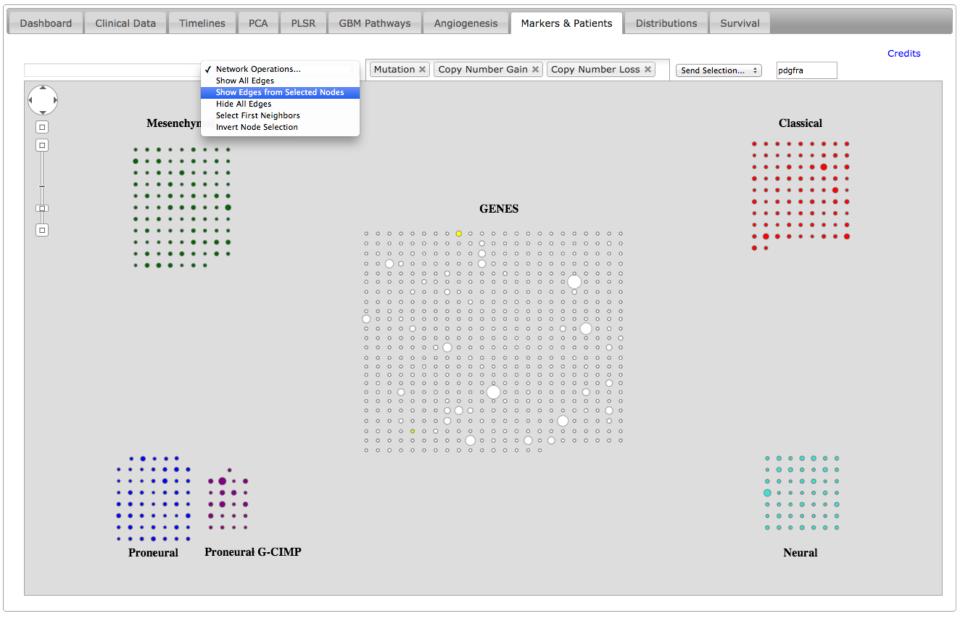


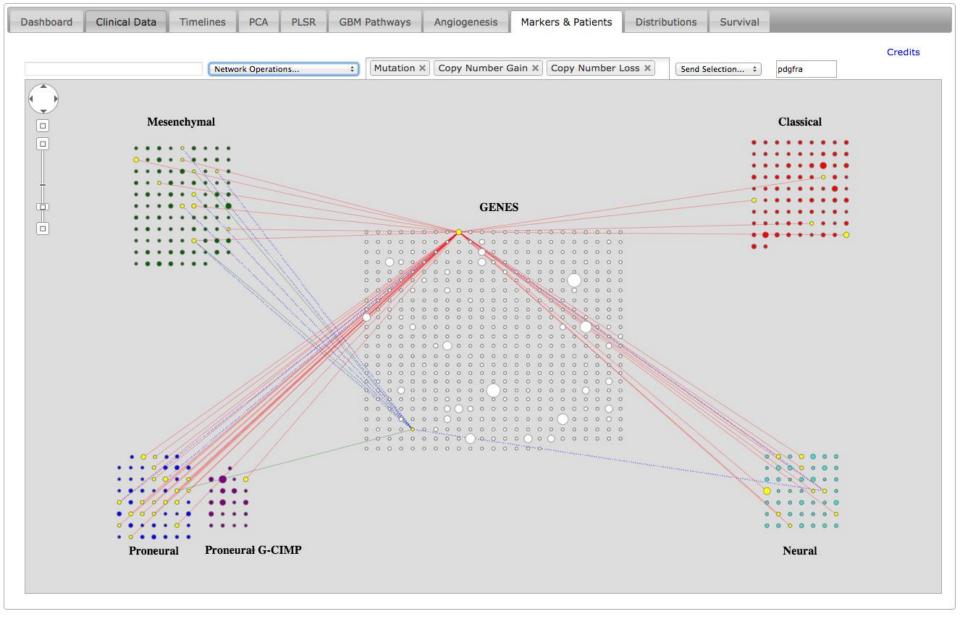


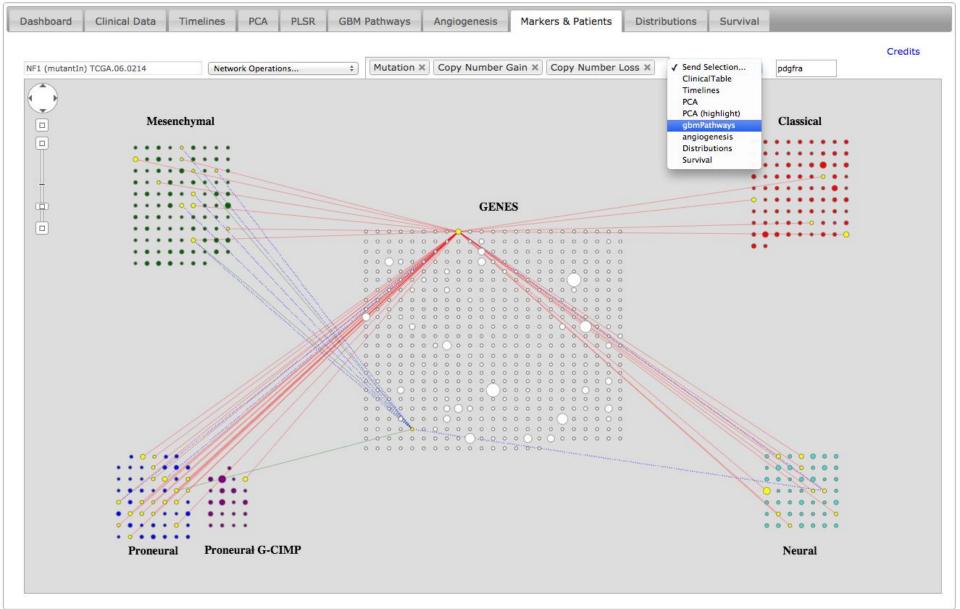


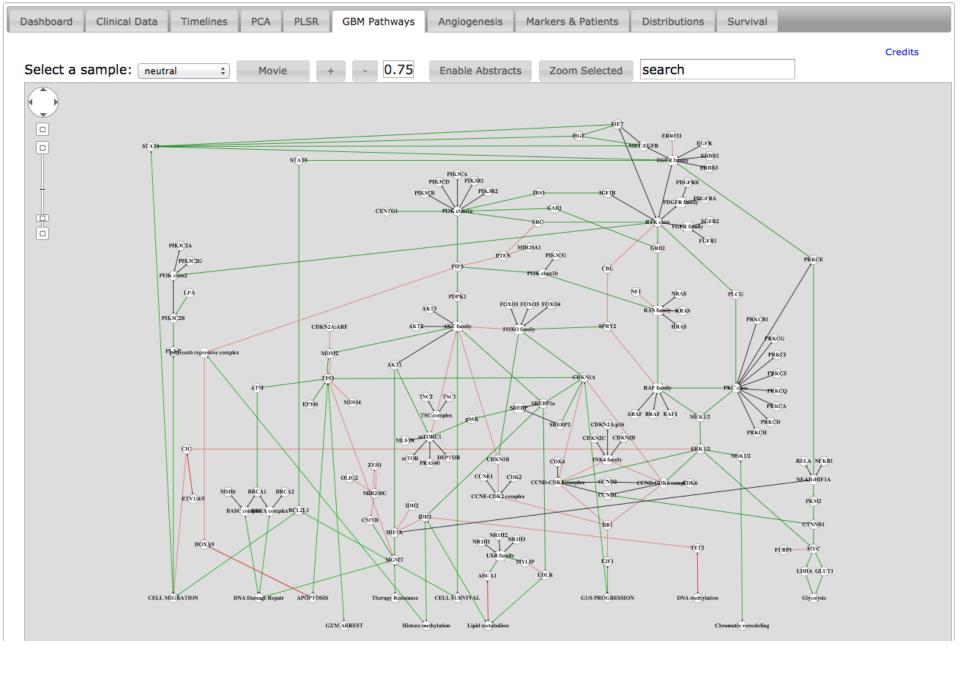


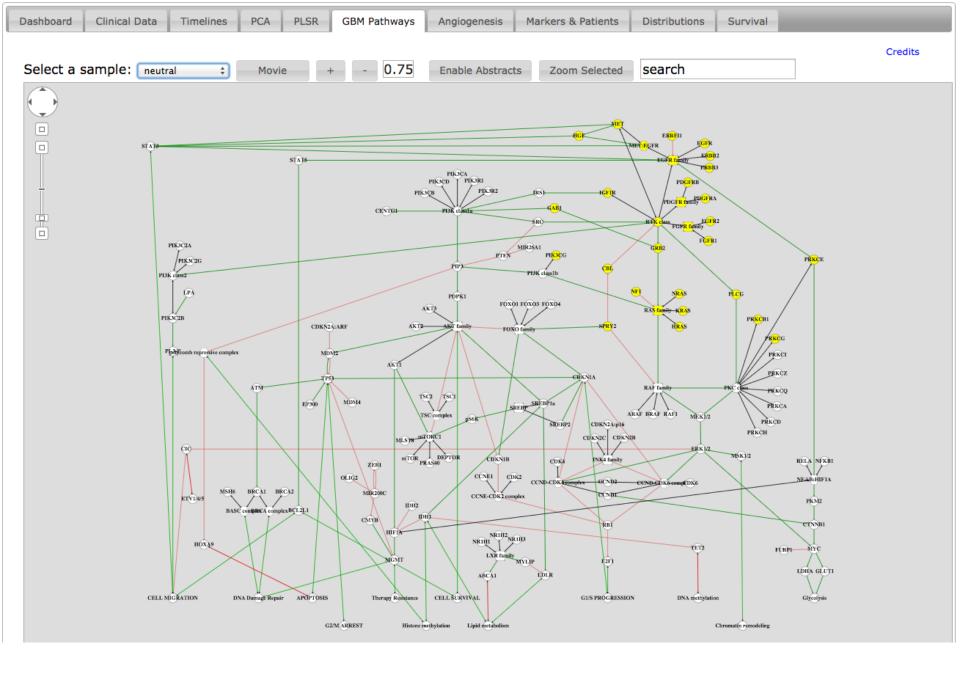


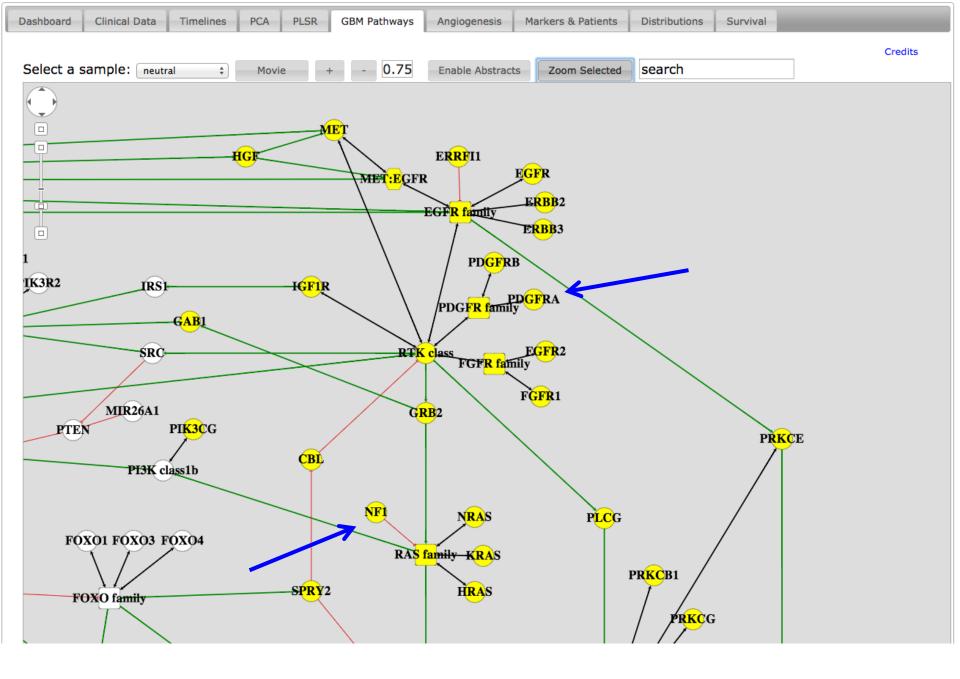


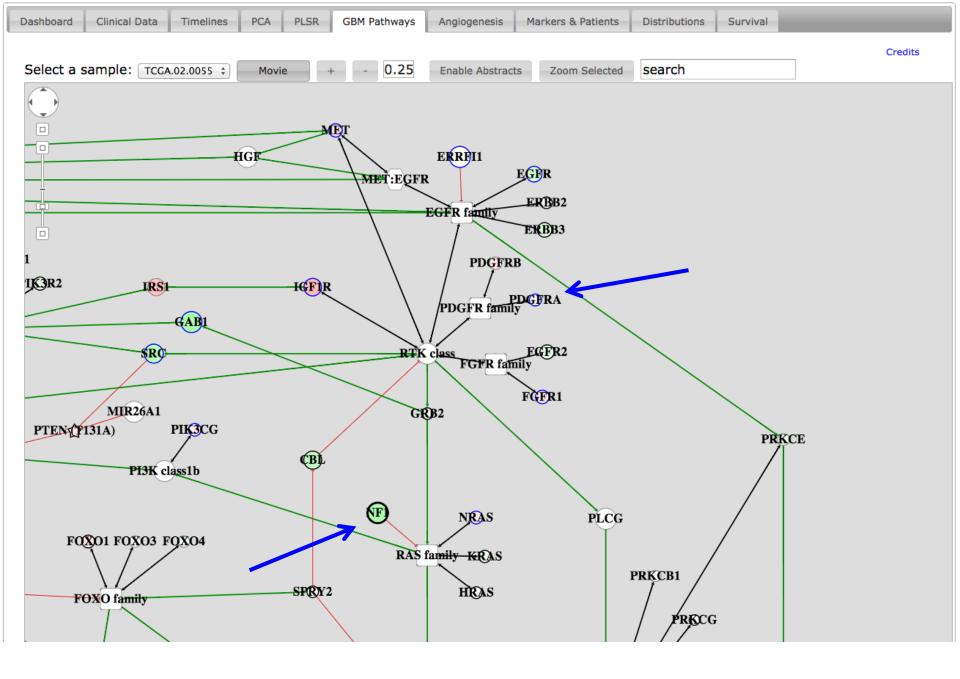


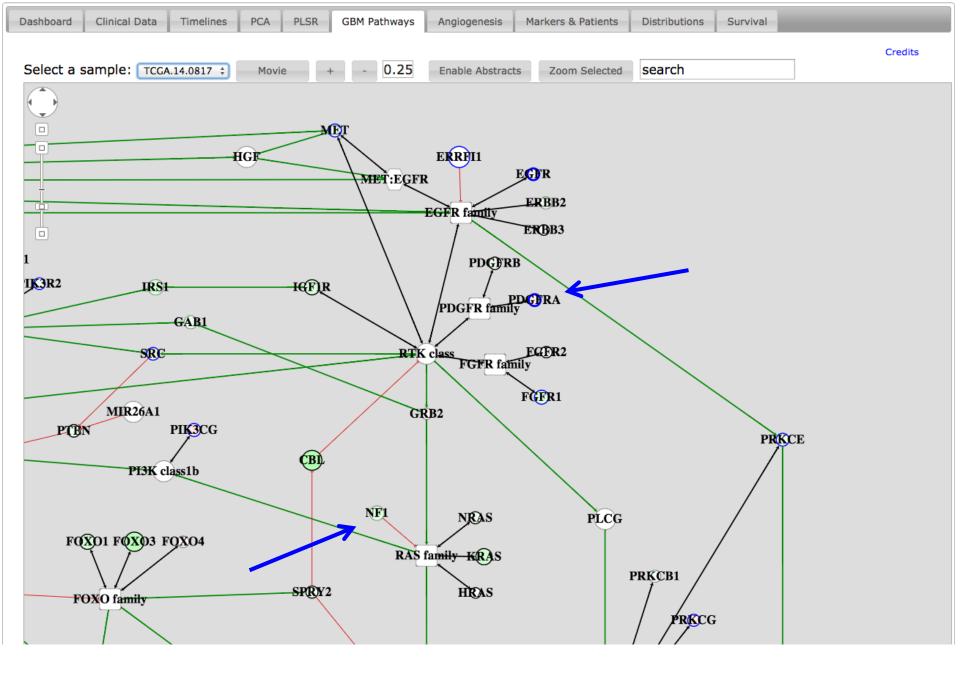


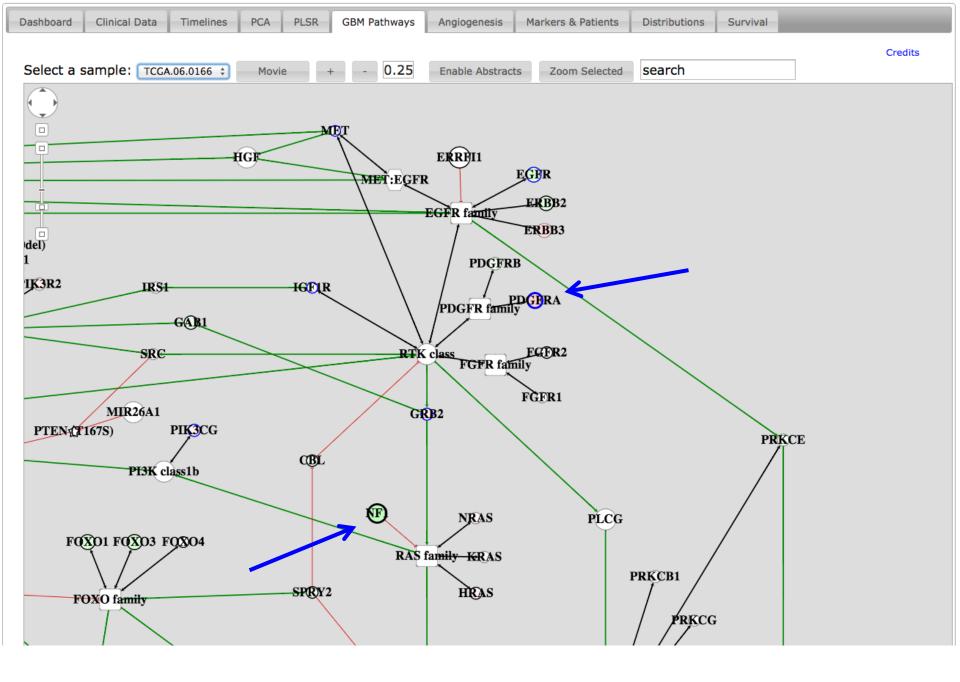


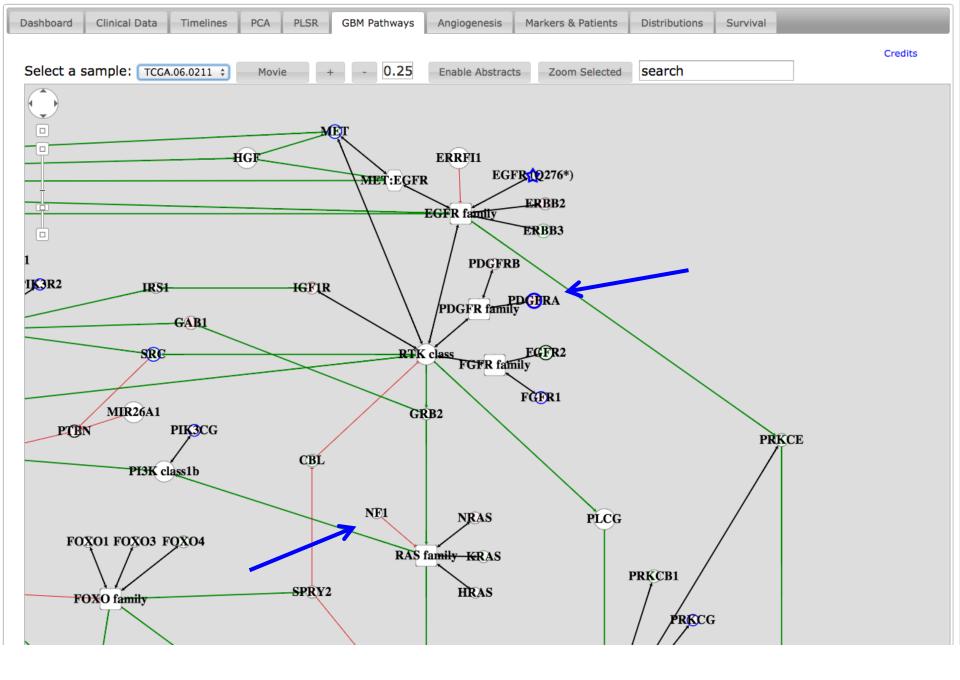


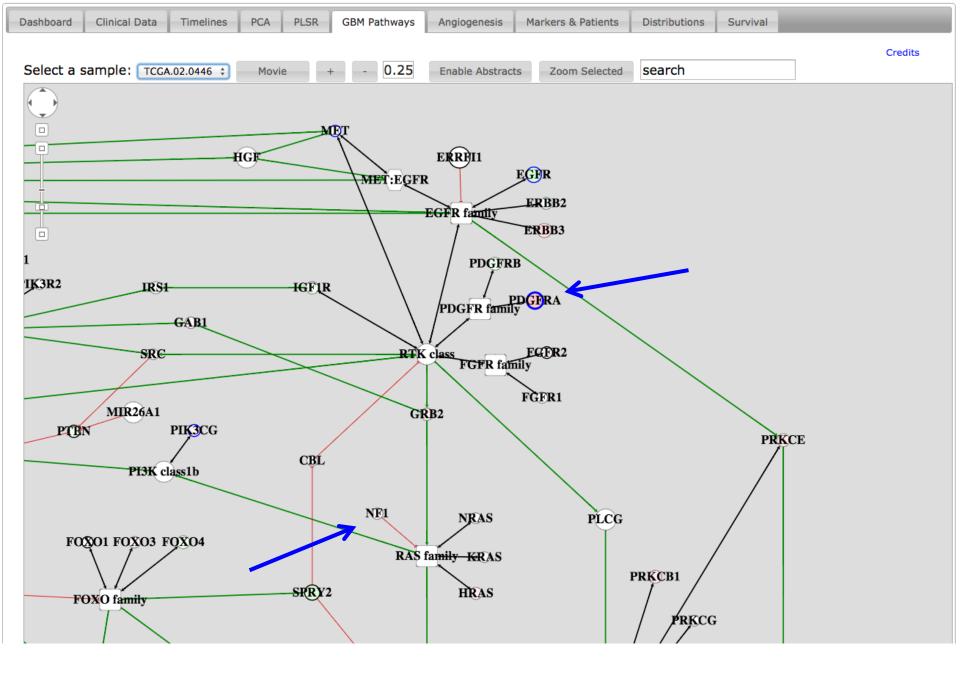


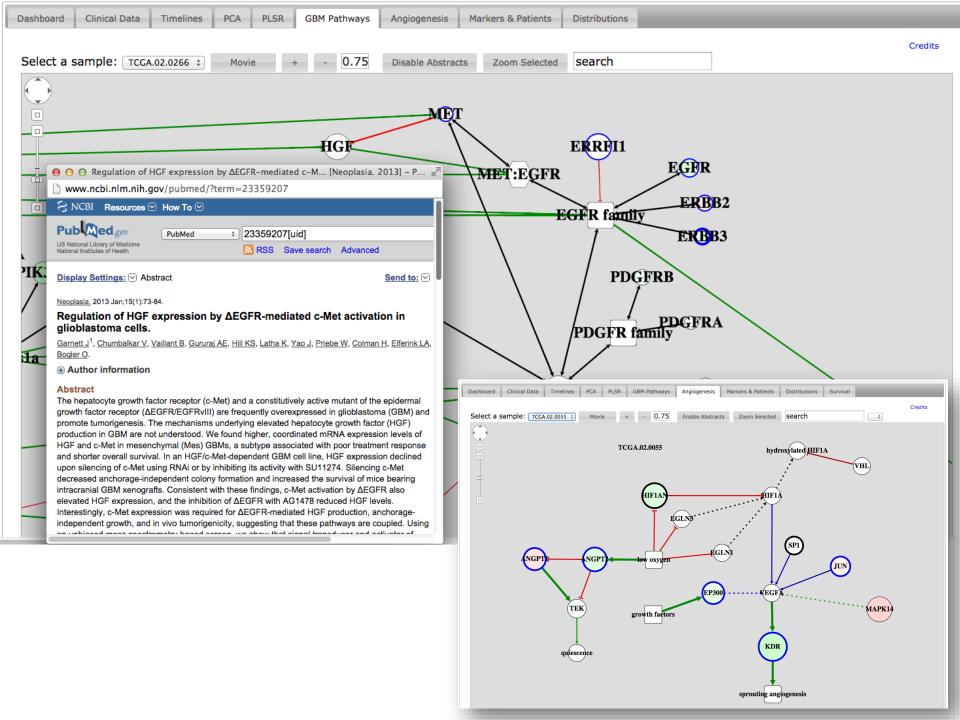


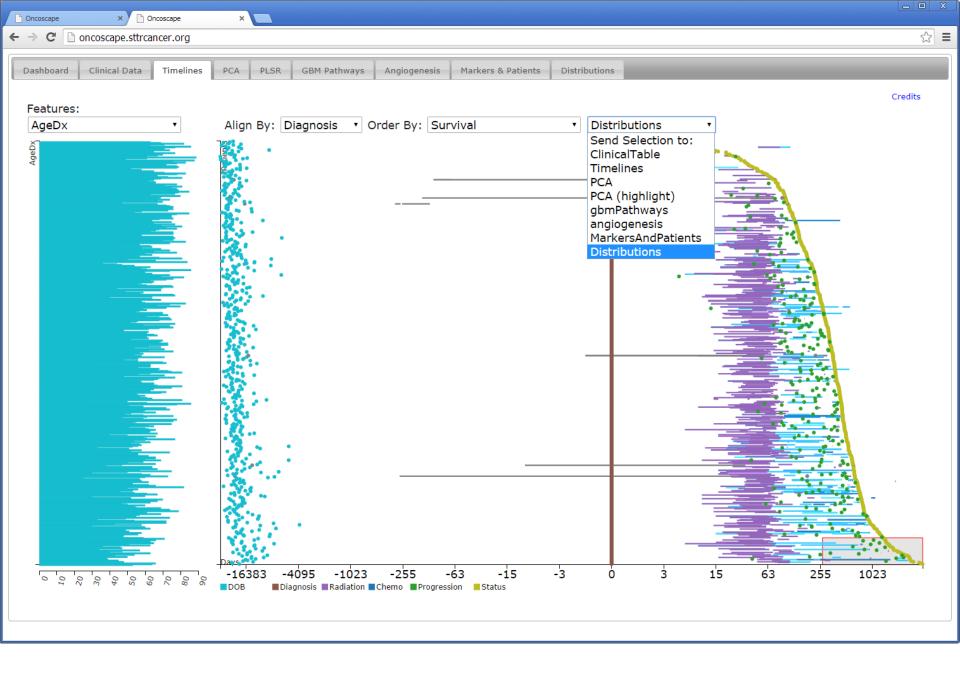


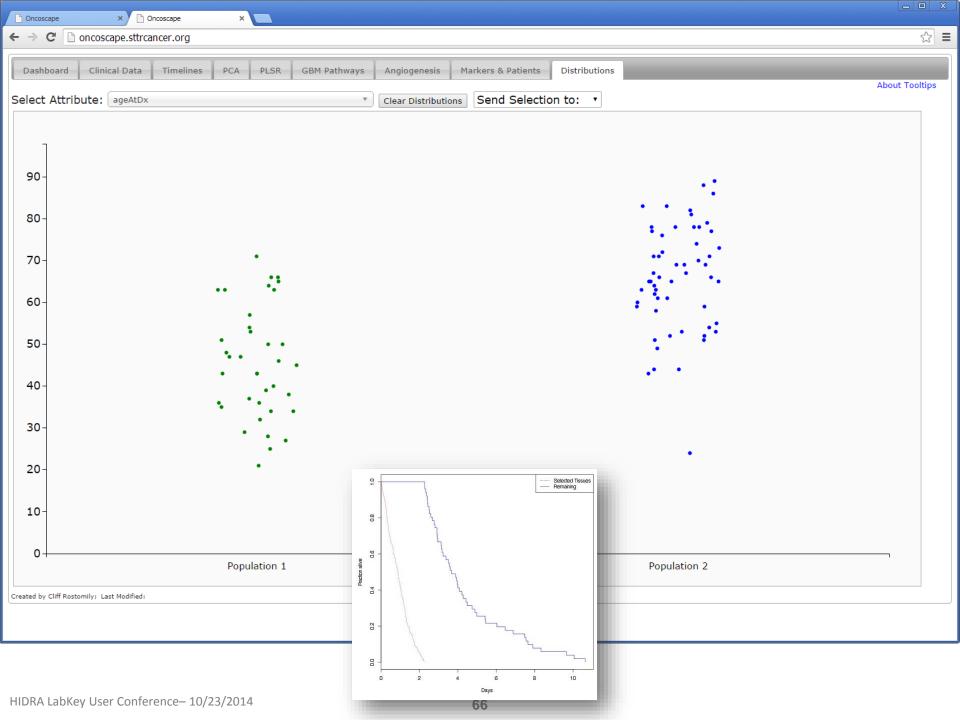


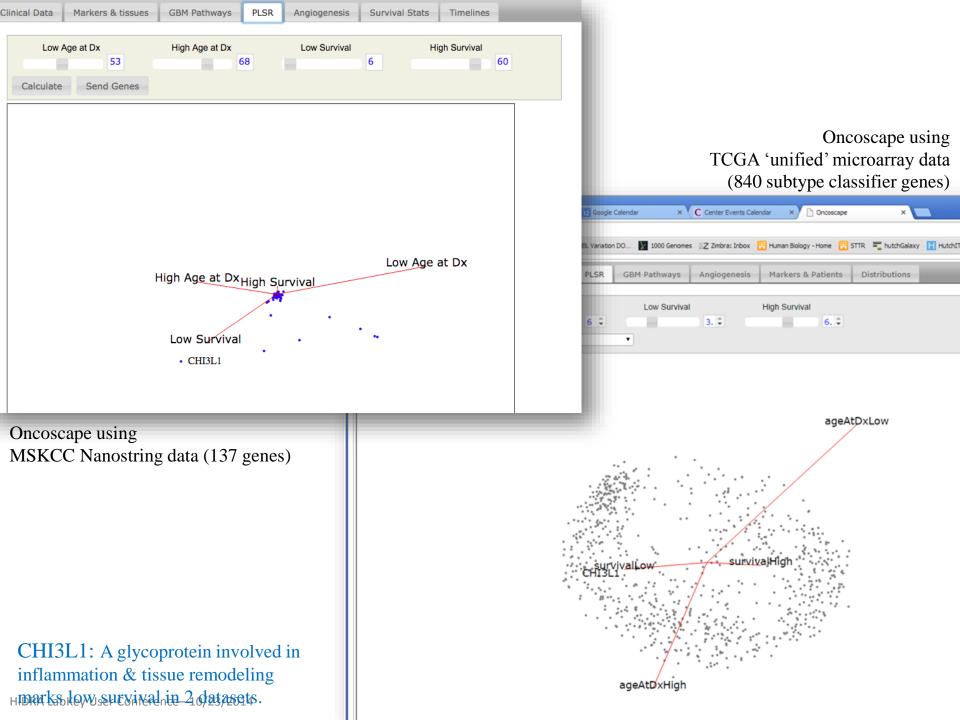












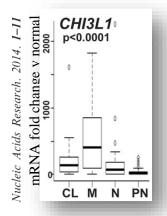
Gene symbol	Gene name	(Cox PH score)	Low risk			High risk		
			Median	Mean	SD*	Median	Mean	SD*
AGT	angiotensinogen	0.03190	-0.801	-2.335	4.366	0.1767	-0.290	2.319
EGFR	epidermal growth factor receptor	-0.05152	1.106	1.261	3.409	2.850	3.105	2.947
CHI3L1	chitinase 3—like 1	0.00442	3.600	3.782	2.965	6.934	6.168	2.624

Table 2. Genes differentially expressed between MSKCC and TCGA long-term survivors versus TCGA patients with survival less than 1 year								
Gene Symbol	P value	Percent Change	Increase or Decrease in LTS vs. patients with survival <1 year	Description				
EFEMP2*	0.000091	31.31	Decrease	EGF containing fibulin-like extracellular matrix protein 2				

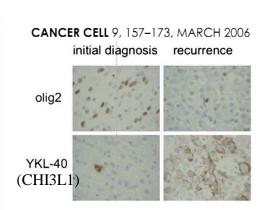
Decrease

0.004373

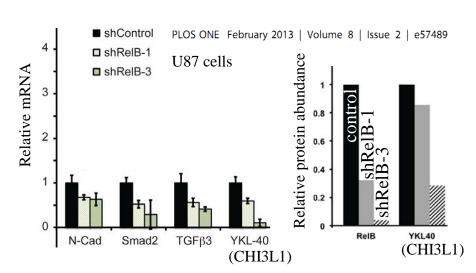
Neuro-Oncology 16(9), 1186-1195, 2014



CHI3L1*



44.95



chitinase 3-like 1 (cartilage glycoprotein-39)

^{*}also significant in REMBRANDT data set.

HIDRA

- Vision and Strategy
- Systems & Requirements
- Program Overview
- Argos user interface
- Next steps

HIDRA 2015

- Multiple disease portals
- NLP pipeline implementation
- Integrate data visualization features from Oncoscape
- Technical and operational rollout of use of PHI and data export

Acknowledgements

- LabKey Software
- UW Medicine ITS and Amalga Team
- SCCA Knowledge Management
- CIT Informatics
- NW BioTrust
- Solid Tumor Translational Research











